

Prácticas ADEVA y Diseño de Experimentos

1. Se dispone de los datos obtenidos por Heyl para medir el valor de la constante gravitatoria universal ' G '. Estas mediciones las realizó utilizando bolas de tres materiales diferentes: oro, platino y vidrio. En la tabla siguiente se presentan los datos ($\times 10^{11} Nm^2/kg^2$)

Oro	Platino	Vidrio
6,683	6,661	6,678
6,681	6,661	6,671
6,676	6,667	6,675
6,678	6,667	6,672
6,679	6,664	6,674
6,672		

- Contrastar la hipótesis de que con cualquiera de los tres materiales se obtiene la misma constante gravitatoria. (Nota.- El valor aceptado para la constante gravitatoria universal es $G = 6,67 \times 10^{-11} Nm^2/kg^2$).
- Calcular un intervalo de confianza para la media y el valor de la desviación típica de la constante gravitatoria obtenida con cada uno de los materiales. (**Heyl.txt**).

2. Se mide la temperatura de una mezcla con cuatro termómetros, obteniéndose los datos siguientes (**Termometros.txt**):

T1	T2	T3	T4
63	64	58	61
63	64	59	61
62	63	59	62
65	64	68	60
66	65		63

- ¿Son los cuatro termómetros análogos?
- Calcule los residuos del modelo anterior y analice si se verifican las hipótesis básicas del modelo.
- Elimine el tercer termómetro y calcule nuevamente la tabla ADEVA. ¿Que conclusiones pueden extraerse?.

3. Un estadístico, preocupado por el diferente resultado que le dan cuatro tipos de pilas diferentes, decide realizar un experimento para comparar la duración por unidad de coste de esas pilas. Para ello compra cuatro pilas de cada tipo (de diferentes lotes), las ordena de modo aleatorio y mide el tiempo durante el cual subministran corriente eléctrica a un

aparato. Los resultados del experimento se proporcionan en la tabla siguiente. Los tipos de pila se denominan respectivamente tipos **1**, **2**, **3** y **4**, la duración se da en minutos, el coste unitario (**CU**) en dólares, la duración por unidad de coste (**DUC**) en minutos por dólar. La variable respuesta es la duración por unidad de coste (**DUC**). Se proporciona también el orden de realización de los experimentos. (**Pilas.txt**)

Tipo	duración	CU	DUC	orden
1	602	0,985	611	1
2	863	0,935	923	2
1	529	0,985	537	3
4	235	0,495	476	4
1	534	0,985	542	5
1	585	0,985	593	6
2	743	0,935	794	7
3	232	0,520	445	8
4	282	0,495	569	9
2	773	0,935	827	10
2	840	0,935	898	11
3	255	0,520	490	12
4	238	0,495	480	13
3	200	0,520	384	14
4	228	0,495	460	15
3	215	0,520	413	16

- Proponga un modelo para estos datos.
- Estudie si existe un tipo de pila que sea el mejor y otro que sea el peor.
- ¿Se cumplen las hipótesis del modelo?

4. Una empresa suministra tubos de escape a la industria del automóvil. En el tubo de escape se coloca un sensor que comunica al ordenador del coche el contenido de CO en los gases de escape. La empresa dispone de dos tipos de sensores A y B basados en principios de medición diferentes. Tanto uno como otro se colocan en el silencioso del tubo de escape. El departamento de I+D de la empresa sospecha que pueden existir diferencias entre las mediciones efectuadas por los dos tipos de sensor y decide realizar un experimento. La primera idea consiste en seleccionar 10 sensores del tipo A y otros 10 del tipo B y colocarlos en los tubos de escape de 20 coches distintos. No obstante un ingeniero de la empresa sugiere que para evitar que la variabilidad entre coches distintos enmascare los resultados del experimento, se utilicen 10 coches y que en cada uno de ellos se monte un sensor A y otro B ambos colocados en el silencioso y en dos posiciones cercanas entre sí.

- ¿Qué procedimiento experimental le parece más adecuado? ¿Por qué?.
- Finalmente se opta por el plan experimental propuesto por el ingeniero. El contenido en partes por millón de CO observado en el experimento fue el que se indica en la tabla (**Sensor.txt**):

A	B	Nº coche
72,1	74,0	1
68,2	68,8	2
70,9	71,2	3
74,3	74,2	4
70,7	71,8	5
66,6	66,4	6
69,5	69,8	7
70,8	71,3	8
68,8	69,3	9
73,3	73,6	10

- ¿Existen diferencias significativas entre los sensores?.
- Dado que el experimento se ha realizado con 10 coches, ¿es preferible utilizar 10 coches del mismo modelo o 10 coches de modelos y cilindradas diferentes?.

5. Se dispone de tres protocolos para medir la tasa de metabolismo basal (en reposo) de las personas, que se denominarán protocolos **1**, **2** y **3**. En el protocolo **1** se ingresa al paciente en el hospital por la tarde, se le da la cena, y reposa en el hospital hasta la mañana siguiente en que se realiza la medición. En el protocolo **2** se da de cenar al paciente en el hospital, se le envía a casa a descansar y vuelve a la mañana siguiente para tomar la medición. En el tercero se indica al paciente que debe cenar en su casa y se le hace ir al hospital a la mañana siguiente para tomar la medición. El objetivo es comprobar si el protocolo no afecta al resultado del análisis, lo cual abarataría costes y permitiría comparar resultados obtenidos en hospitales que usasen diferentes protocolos. Dado que se sabe de la gran variabilidad entre pacientes se toman nueve pacientes a los que se aplican los tres protocolos en días diferentes. Los resultados obtenidos se proporcionan en la tabla siguiente (**Metabolismo.txt**):

	Protocolo		
	1	2	3
Paciente 1	7131	6846	7095
Paciente 2	8062	8573	8685
Paciente 3	6921	7287	7132
Paciente 4	7249	7554	7471
Paciente 5	9551	8866	8840
Paciente 6	7046	7681	6939
Paciente 7	7715	7535	7831
Paciente 8	9862	10087	9711
Paciente 9	7812	7708	8179

- Proponga un modelo para los datos.
- ¿Producen diferentes resultados los tres protocolos?

- ¿Hay algún paciente cuya tasa de metabolismo basal sea superior al resto?
- Verifique las hipótesis del modelo.

6. Cierta industria de lentes para gafas desea comparar dos tipos de recubrimiento antireflectante. Los dos tipos tienen idéntico aspecto y prestaciones, pero antes de decidirse por uno u otro desean comprobar si el tipo de recubrimiento influye en el deterioro que sufre la lente. En un principio, piensan seleccionar al azar 20 personas que utilicen gafas, y asignar aleatoriamente gafas sometidas al tratamiento A a 10 de ellas y gafas con el tratamiento B a las otras 10. Al cabo de 6 meses se realizaría un control y tras tomar medidas con el instrumento adecuado, se analizaría si existían o no diferencias estadísticamente significativas.

- ¿Qué inconveniente tiene el método planteado? ¿Es posible eliminar el efecto de las personas en el desgaste de la lente?.
- Se decide construir gafas con una lente de cada tipo. Si alguien las desgasta lo hará con ambos tratamientos, igual que si las desgasta poco. Al cabo de seis meses se mide el desgaste y se obtienen los valores que se indican en la tabla. ¿Qué lente elegiríamos?. Nota: La letra entre paréntesis indica la posición (izquierda o derecha) a que se ha asignado aleatoriamente cada lente (**Gafas.txt**).

Individuo	Lente A	Lente B
1	6.7(I)	6.9(D)
2	5.0(I)	5.8(D)
3	3.6(D)	4.1(I)
4	6.2(I)	7.0(D)
5	5.9(D)	7.0(I)
6	4.0(D)	4.6(I)
7	5.2(D)	5.5(I)
8	4.5(I)	5.0(D)
9	4.4(D)	4.3(I)
10	4.1(I)	4.8(D)

7. En 1982 se llevó a cabo un experimento para estudiar las estrategias de reducción de precio y de colocación de productos en un supermercado para aumentar ventas. Para ello se consideró el factor precio a tres niveles (Usual, Reducido y Con Pérdidas para el supermercado) y el factor colocación a otros tres niveles (espacio normal, espacio normal + fin de pasillo y espacio doble). La variable respuesta estudiada fue el número de unidades vendidas de un cierto producto en una semana. Para cada tratamiento se tomó la medida en tres supermercados de características similares. Los datos se muestran en la tabla siguiente (**Supermercado.txt**):

Estrategia de colocación	Precio		
	Usual	Reducido	Con pérdidas
Normal	989-1025-1030	1211-1215-1182	1577-1559-1598
Normal + fin de pasillo	1191-1233-1221	1860-1910-1926	2492-2527-2511
Doble	1226-1202-1180	1516-1501-1498	1801-1833-1852

- Identifique el modelo de que se trata, estime sus parámetros y realice la diagnosis.
- Indique si las interacción es significativa o no e interprétela si lo fuere.
- ¿Cuál sería la estrategia óptima para vaciar el almacén de un cierto producto?

8. Cierta Organismo Público (O.P.) encargado de certificar la composición de aleaciones de metales preciosos, debe seleccionar entre dos Laboratorios al más capacitado para la realización de futuros análisis de gran precisión. Para tomar la decisión los somete a la siguiente prueba: Prepara tres aleaciones A , B y C que contienen proporciones distintas de oro. De cada una de ellas envía cuatro muestras a cada uno de los dos laboratorios. Así pues, cada laboratorio recibe un lote de 12 muestras (codificadas) ordenadas aleatoriamente sin conocer como han sido obtenidas. Los resultados recibidos por el O.P. se muestran en la tabla siguiente (entre paréntesis las medias de las casillas) (**Laboratorios.txt**):

	<i>Aleac. A</i>		<i>Aleac. B</i>		<i>Aleac. C</i>	
Lab. I	10,96	11,03	10,95	11,00	11,07	11,01
	11,08	11,01	11,04	10,97	10,97	11,03
	<i>(11,02)</i>		<i>(10,99)</i>		<i>(11,02)</i>	
Lab. II	10,97	10,96	10,97	10,96	11,02	11,00
	10,94	10,95	10,97	10,98	11,01	11,01
	<i>(10,955)</i>		<i>(10,97)</i>		<i>(11,01)</i>	

- Determinar si existen diferencias entre los resultados de los laboratorios y si éstos han encontrado diferencias entre las aleaciones.
- Indicar si podemos aceptar que se verifican las hipótesis del modelo y en caso negativo qué medidas se deben adoptar para analizar los datos.
- Contrastar si las varianzas de los dos laboratorios son iguales, sabiendo que las tres aleaciones tienen composición distinta. Interpretar el resultado.
- El O.P. conoce exáctamente el porcentaje en oro de la aleación A (11 %), de la B (11,02 %) y de la C (11,04 %). Con esta información comparar los resultados de los laboratorios.

9. Se ha realizado un experimento para estudiar las fuentes de variabilidad de la resistencia a la compresión de cemento tipo Portland. El cemento ha sido mezclado con agua por tres obreros

diferentes (mezcladores) durante un tiempo fijo. Después, la resistencia de las probetas generadas ha sido medida por otros tres obreros diferentes (medidores). Cada mezclador ha generado doce probetas, que se han dividido en tres grupos de cuatro; cada uno de esos grupos de cuatro ha sido asignado a un medidor. Los datos obtenidos para la resistencia a la compresión de cada probeta, dados en libras por pulgada cuadrada, se proporcionan en la tabla siguiente (**Cemento.txt**):

	Medidor 1	Medidor 2	Medidor 3
Mezclador 1	5280	4340	4160
	5520	4400	5180
	4760	5020	5320
	5800	6200	4600
Mezclador 2	4420	5340	4180
	5280	4880	4800
	5580	4960	4600
	4900	6200	4480
Mezclador 3	5360	5720	4460
	6160	4760	4930
	5680	5620	4680
	5500	5560	5600

- ¿Existen diferencias entre las resistencias dadas por los diferentes medidores? ¿Y entre las probetas generadas por cada mezclador?
- ¿Es significativa, con nivel de significación del 5%, la interacción entre medidores y mezcladores?
- ¿Se cumplen las hipótesis del modelo?

10. Se ha realizado un diseño experimental para determinar la influencia de dos factores, combinación de hidrocarburos y cantidad de hidrógeno, en el rendimiento de un proceso químico complejo. Se estudiaron cuatro combinaciones de hidrocarburo (A, B, C y D) y tres niveles en el contenido de hidrógeno (1, 2 y 3). En cada tratamiento se realizaron cuatro réplicas. En la tabla se presentan los resultados: mejora en tanto por mil respecto al procedimiento estándar (**Hidrogeno.txt**).

- Compare las medias de los cuatro niveles del factor *Hidrocarburo* y las de los tres niveles del factor *Hidrógeno*. Indique si existen diferencias significativas con nivel de significación 0,05.
- Elija el tratamiento que proporciona el rendimiento óptimo, justificando la respuesta. De un intervalo de confianza para el valor medio en dichas condiciones con nivel de confianza del 95%.

	A	B	C	D	Etapa
	10.3	10.5	7.2	13.0	1
	11.1	8.2	5.3	12.9	1
1	15.3	9.7	12.5	5.3	2
	2.1	8.9	19.1	12.0	2
	25.8	20.6	29.7	17.6	1
	25.7	17.1	26.3	12.0	1
2	28.9	21.4	22.4	24.6	2
	27.8	17.3	25.9	23.1	2
	28.5	21.0	30.4	20.5	1
	31.2	26.8	26.6	26.2	1
3	24.8	19.4	34.4	27.8	2
	26.5	22.2	27.5	21.9	2

11. Unos alumnos de la universidad de Tufts (Massachussets, E.U.A.), preocupados por el estado de corrosión de las tuberías de su universidad, decidieron realizar el siguiente experimento. Tomaron muestras de agua corriente haciendo variar los factores Campus, Tipo de edificio y Antigüedad del edificio. Se midió la concentración de hierro en el agua corriente (mg/dm^3) y para cada posible combinación de factores se tomaron dos observaciones. Los datos se muestran en la siguiente tabla (**Tuffs.txt**):.

Factor			Concentración de Fe	
Antigüedad	Tipo	Campus		
Viejo	Académico	Medford	0,23	0,28
Nuevo	Académico	Medford	0,36	0,29
Viejo	Residencial	Medford	0,03	0,06
Nuevo	Residencial	Medford	0,05	0,02
Viejo	Académico	Somerville	0,08	0,05
Nuevo	Académico	Somerville	0,03	0,08
Viejo	Residencial	Somerville	0,04	0,07
Nuevo	Residencial	Somerville	0,02	0,06

- Identifique el modelo de que se trata, estime sus parámetros y realice la diagnosis.
- Si no se cumplieran las hipótesis del modelo indique qué podría hacerse para remediarlo.
- Estudie las interacciones e interprete las que resulten significativas.

12. En una investigación sobre el efectos de cuatro fertilizantes (A, B, C y D) en el rendimiento de distintas semillas de trigo (4 tipos de semillas) en cuatro calidades distintas de terreno se obtuvieron los siguientes resultados del rendimiento en % (**Semillas.txt**):

Semilla	Terreno			
	T1	T2	T3	T4
S1	35,5(A)	24,5(B)	14,7(C)	35,5(D)
S2	14,3(B)	6,2(C)	13,7(D)	24,5(A)
S3	14,1(C)	16,2(D)	34,3(A)	19,7(B)
S4	15,0(D)	64,5(A)	34,6(B)	19,0(C)

- ¿Se puede concluir que hay diferencias entre las semillas?, ¿y entre los fertilizantes?. ¿Afecta la calidad del terreno al rendimiento?
- Recomendar cuál es el tipo de semilla y el fertilizante con el criterio del mayor rendimiento.

13. Un proceso físico-químico requiere la explosión controlada de ciertas sustancias en un reactor. Se ha construido un reactor experimental con la finalidad de medir las presiones máximas que debe soportar la cámara del mismo en el momento de la explosión. El factor que más influye en la presión es la temperatura de operación y se desea diseñar el reactor para que trabaje a temperaturas extremas. Se ha realizado un experimento que contempla tres temperaturas: Baja (-40°), Media (20°) y Alta (60°). Además se estudiaron otros dos factores: tipo de explosivo utilizado (A, B, C y D) y emplazamiento (I y II). En cada uno de los 24 tratamientos se repitió el experimento 6 veces. La variable respuesta es la presión máxima alcanzada en el reactor. Los datos están en el fichero **Reactor.txt**

- Realiza el análisis de la varianza considerando los tres factores e interpreta los resultados.
- El procedimiento de seguridad del proceso indica que en caso de no cumplir la hipótesis de homocedasticidad se repita el análisis de los datos transformados mediante la función logarítmica. Interpreta los resultados. ¿Es adecuada esta recomendación? Justifica tu decisión.
- Si la transformación logarítmica no es adecuada, el procedimiento recomienda analizar el experimento para cada una de las tres temperaturas por separado. Realiza el análisis en este caso e interpreta los resultados.
- El objetivo del experimento es establecer la presión media esperada y la desviación típica en el peor de los tratamientos, considerando como peor el que proporciona una mayor presión en el reactor. Calcula un intervalo de confianza para la media y el valor de la desviación típica de las presiones en las condiciones solicitadas.
- ¿Cuál es la temperatura que proporciona más presión? ¿Qué presiones son de esperar en un reactor trabajando a la temperatura anterior?.

PRÁCTICAS DE REGRESIÓN

1. En el negocio de la madera es de interés ser capaz de estimar la cantidad de madera disponible en un bosque. Es laborioso medir el volumen de madera de un árbol, sin embargo, es sencillo medir su altura y su diámetro en la base. Resulta pues interesante para el maderero disponer de una relación empírica que relacione estas dos últimas magnitudes con el volumen del árbol. Con este fin se midieron el diámetro a 4,5 pies de altura (en pies), la altura (en pies) y el volumen (en pies cúbicos) de 31 cerezos negros del Bosque Nacional del Allegheny (Pensilvania, USA). En la siguiente tabla se muestran los datos.

Diámetro	8,3	8,6	8,8	10,5	10,7	10,8	11,0	11,0	11,1	11,2	11,3
Altura	70	65	63	72	81	83	66	75	80	75	79
Volumen	10,3	10,3	10,2	16,4	18,8	19,7	15,6	18,2	22,6	19,9	24,2
Diámetro	11,4	11,4	11,7	12,0	12,9	12,9	13,3	13,7	13,8	14,0	14,2
Altura	76	76	69	75	74	85	86	71	64	78	80
Volumen	21,0	21,4	21,3	19,1	22,2	33,8	27,4	25,7	24,9	34,5	31,7
Diámetro	14,5	16,0	16,3	17,3	17,5	17,9	18,0	18,0	20,6		
Altura	74	72	77	81	82	80	80	80	87		
Volumen	36,3	38,3	42,6	55,4	55,7	58,3	51,5	51,0	77,0		

a. Estime la relación entre el volumen y el diámetro y altura de estos árboles.

2. En 1976 los investigadores Allison y Cicchetti realizaron un estudio sobre una serie de especies de mamíferos. Entre los datos que recopilaron se encuentran el peso medio del **cuerpo** (dado en kilogramos) y el peso medio del **cerebro** (dado en gramos) de 66 de esas especies. En la siguiente tabla se reproducen los datos recopilados. Se trata de estudiar la posible relación entre estas dos variables.

Especie	Cuerpo	Cerebro
Zorro del Ártico	3,385	44,50
Mono buho	0,480	15,50
Castor de montaña	1,350	8,10
Vaca	465,000	423,00
Lobo gris	36,330	119,50
Cabra	27,660	115,00
Corzo	14,830	98,20
Conejo de Indias	1,040	5,50
Mono verde	4,190	58,00
Chinchilla	0,425	6,40
Ardilla	0,101	4,00
Ardilla del Artico	0,920	5,70
Rata Gigante africana	1,000	6,60

Especie	Cuerpo	Cerebro
Musaraña	0,005	0,14
Topo	0,060	1,00
Armadillo	3,500	10,80
Daman arboreo	2,000	12,30
Opossum	1,700	6,30
Elefante asiático	2547,000	4603,00
Murciélago	0,023	0,30
Burro	187,100	419,00
Caballo	521,000	655,00
Erizo	0,785	3,50
Mono	10,000	115,00
Gato	3,300	25,60
Galago	0,200	5,00
Gineta	1,410	17,50
Girafa	529,000	680,00
Gorila	207,000	406,00
Foca gris	85,000	325,00
Daman de los matorrales	0,750	12,30
Hombre	62,000	1320,00
Elefante africano	6654,000	5712,00
Opossum acuático	3,500	3,90
Mono Rhesus	6,800	179,00
Canguro	35,000	56,00
Marmota	4,050	17,00
Hamster	0,120	1,00
Ratón	0,023	0,40
Murcielago pardo	0,010	0,25
Loris	1,400	12,50
Okapi	250,000	490,00
Conejo	2,500	12,10
Oveja	55,500	175,00
Jaguar	100,000	157,00
Chimpance	52,160	440,00
Babuino	10,550	179,50
Erizo	0,550	2,40
Armadillo gigante	60,000	81,00
Daman de El Cabo	3,600	21,00
Mapache	4,288	39,20
Rata	0,280	1,90
Topo americano	0,075	1,20
Rata topo	0,122	3,00
Musaraña almizclera	0,048	0,33

Especie	Cuerpo	Cerebro
Cerdo	192,000	180,000
Equidna	3,000	25,000
Tapir brasileño	160,000	169,000
Tenrec	0,900	2,600
Falangero	1,620	11,400
Musaraña arborícola	0,104	2,500
Zorro rojo	4,235	50,400

- Construya el modelo de regresión simple tomando como variable dependiente el peso del cerebro y como regresor el peso del cuerpo.
- Facilite el coeficiente de determinación de la regresión, la estimación de la pendiente y un intervalo de confianza del 95% para ésta.
- ¿Qué observaciones identifica como influyentes y cuáles identifica como generadoras de residuos muy grandes? Compruebe ésto sobre los pertinentes gráficos de residuos.
- Tome logaritmos de ambas variables y proporcione el nuevo valor de R^2 .
- Revise los gráficos de residuos. ¿Puede inferirse de ellos que no se cumple alguna de las hipótesis del modelo?

3. Se ha medido el contenido medio en **alquitrán** (en miligramos), **nicotina** (en miligramos) y **monóxido de carbono** producido en la combustión (en miligramos) de cigarrillos de una serie de marcas de tabaco. Los datos se proporcionan en la siguiente tabla. Se pretende estudiar la relación entre el monóxido de carbono producido en la combustión de un cigarrillo y su contenido en alquitrán y nicotina.

Marca	Alquitrán	Nicotina	CO
Alpine	14,1	0,86	13,6
Benson & Edges	16,0	1,06	16,6
Bull Durham	29,8	2,03	23,5
Camel lights	8,0	0,67	10,2
Carlton	4,1	0,40	5,4
Chesterfield	15,0	1,04	15,0
Golden lights	8,8	0,76	9,0
Kent	12,4	0,95	12,3
Kool	16,6	1,12	16,3
L & M	14,9	1,02	15,4
Lark lights	13,7	1,01	13,0
Marlboro	15,1	0,90	14,4
Merit	7,8	0,57	10,0
Multi Filter	11,4	0,78	10,2
Newport light	9,0	0,74	9,5
Now	1,0	0,13	1,5
Old Gold	17,0	1,26	18,5
Pall Mall lig	12,8	1,08	12,6
Raleigh	15,8	0,96	17,5
Salem Ultra	4,5	0,42	4,9
Tareyton	14,5	1,01	15,9
True	7,3	0,61	8,5
Viceroy Rich	8,6	0,69	10,6
Virginia Slim	15,2	1,02	13,9
Winston light	12,0	0,82	14,9

- Estudie la relación entre el monóxido de carbono y el alquitrán.
- Estudie la relación entre el monóxido de carbono y la nicotina.
- Realice la regresión múltiple del monóxido de carbono frente a la nicotina y el alquitrán simultáneamente. ¿Qué diferencias llamativas se aprecia entre este estudio y los dos estudios previos? ¿A qué pueden deberse?

4. Durante las décadas de los cuarenta y cincuenta del siglo XIX, el físico escocés J.D. Forbes realizó estudios con los que pretendía determinar la altitud sobre el nivel del mar de cumbres montañosas. Para realizar este estudio, Forbes tomó medidas de la presión y la temperatura de ebullición del agua en diferentes ubicaciones de los Alpes y Escocia. Las medidas de presión fueron registradas en pulgadas de mercurio. Las medidas de temperatura las registró en grados Fahrenheit. El procedimiento habitual para determinar la altitud de un punto en una cordillera consistía en medir la presión atmosférica en dicha posición y utilizar la relación conocida entre la presión y la altitud. El inconveniente de este procedimiento residía en la dificultad que acarrearba el

transporte por lugares montañosos de los frágiles barómetros de la época. Forbes decidió calcular la presión barométrica a partir de la temperatura de ebullición del agua, que es sencilla de medir. Los datos obtenidos por Forbes son los siguientes:

Observación	Temperatura	Presión
1	194,5	20,79
2	194,3	20,79
3	197,9	22,40
4	198,4	22,67
5	199,4	23,15
6	199,9	23,35
7	200,9	23,89
8	201,1	23,99
9	201,4	24,02
10	201,3	24,01
11	203,6	25,14
12	204,6	26,57
13	209,5	28,49
14	208,6	27,76
15	210,7	29,04
16	211,9	29,88
17	212,2	30,06

- a. Estime el modelo de regresión que relacione la variable presión en función de la temperatura de ebullición del agua.

5. Se ha recopilado un conjunto de datos relacionados con 155 modelos de vehículos salidos al mercado americano entre los años 1978 y 1982. En la tabla siguiente se proporcionan el consumo (en millas por galón), el peso (en kilogramos), la potencia (en caballos de vapor), la aceleración, el precio (en dólares), el año de salida al mercado y el origen (valor 1 para modelos americanos, 2 para europeos y 3 para japoneses) de estos modelos.

vehículo	consumo	potencia	año	peso	precio	origen	aceleración
1	43.1	48	78	1985	2400	2	21,5
2	36.1	66	78	1800	1900	1	14,4
3	32.8	52	78	1985	2200	3	19,4
4	39.4	70	78	2070	2725	3	18,6
5	36.1	60	78	1800	2250	3	16,4
6	19.9	110	78	3365	3300	1	15,5
7	19.4	140	78	3735	3125	1	13,2
8	20.2	139	78	3570	2850	1	12,8
9	19.2	105	78	3535	2800	1	19,2
10	20.5	95	78	3155	3275	1	18,2
11	20.2	85	78	2965	2375	1	15,8
12	25.1	88	78	2720	2275	1	15,4
13	20.5	100	78	3430	2700	1	17,2
14	19.4	90	78	3210	2300	1	17,2
15	20.6	105	78	3380	3300	1	15,8
16	20.8	85	78	3070	2425	1	16,7
17	18.6	110	78	3620	2700	1	18,7
18	18.1	120	78	3410	2425	1	15,1
19	19.2	145	78	3425	3900	1	13,2
20	17.7	165	78	3445	4400	1	13,4
21	18.1	139	78	3205	2525	1	11,2
22	17.5	140	78	4080	3000	1	13,7
23	30	68	78	2155	2100	1	16,5
24	27.5	95	78	2560	2975	3	14,2
25	27.2	97	78	2300	2775	3	14,7
26	30.9	75	78	2230	2250	1	14,5
27	21.1	95	78	2515	3700	3	14,8
28	23.2	105	78	2745	3200	1	16,7
29	23.8	85	78	2855	2400	1	17,6
30	23.9	97	78	2405	2975	3	14,9
31	20.3	103	78	2830	4475	2	15,9
32	17	125	78	3140	5875	2	13,6
33	21.6	115	78	2795	4200	2	15,7
34	16.2	133	78	3410	5450	2	15,8
35	31.5	71	78	1990	3675	2	14,9
36	29.5	68	78	2135	3425	3	16,6
37	21.5	115	79	3245	3925	1	15,4
38	19.8	85	79	2990	3200	1	18,2
39	22.3	88	79	2890	2975	1	17,3
40	20.2	90	79	3265	3150	1	18,2

vehículo	consumo	potencia	año	peso	precio	origen	aceleracion
41	20.6	110	79	3360	3325	1	16,6
42	17	130	79	3840	4650	1	15,4
43	17.6	129	79	3725	4850	1	13,4
44	16.5	138	79	3955	5725	1	13,2
45	18.2	135	79	3830	4025	1	15,2
46	16.9	155	79	4360	5225	1	14,9
47	15.5	142	79	4054	4825	1	14,3
48	19.2	125	79	3605	4100	1	15
49	18.5	150	79	3940	4725	1	13
50	31.9	71	79	1925	3100	2	14
51	34.1	65	79	1975	2750	3	15,2
52	35.7	80	79	1915	2700	1	14,4
53	27.4	80	79	2670	2725	1	15
54	25.4	77	79	3530	15475	2	20,1
55	23	125	79	3900	9900	1	17,4
56	27.2	71	79	3190	4675	2	24,8
57	23.9	90	79	3420	4050	1	22,2
58	34.2	70	79	2200	2625	1	13,2
59	34.5	70	79	2150	2775	1	14,9
60	31.8	65	79	2020	2750	3	19,2
61	37.3	69	79	2130	2275	2	14,7
62	28.4	90	79	2670	3525	1	16
63	28.8	115	79	2595	3625	1	11,3
64	26.8	115	79	2700	3525	1	12,9
65	33.5	90	79	2556	3325	1	13,2
66	41.5	76	80	2144	7000	2	14,7
67	38.1	60	80	1968	3850	3	18,8
68	32.1	70	80	2120	2900	1	15,5
69	37.2	65	80	2019	3525	3	16,4
70	28	90	80	2678	3625	1	16,5
71	26.4	88	80	2870	3525	1	18,1
72	24.3	90	80	3003	3625	1	20,1
73	19.1	90	80	3381	3700	1	18,7
74	34.3	78	80	2188	5900	2	15,8
75	29.8	90	80	2711	5500	3	15,5
76	31.3	75	80	2542	4675	3	17,5
77	37	92	80	2434	4050	3	15
78	32.2	75	80	2265	3975	3	15,2
78	46.6	65	80	2110	3350	3	17,9
80	27.9	105	80	2800	3200	1	14,4

vehículo	consumo	potencia	año	peso	precio	origen	aceleración
81	40.8	65	80	2110	3300	3	19,2
82	44.3	48	80	2085	3900	2	21,7
83	43.4	48	80	2335	3825	2	23,7
84	36.4	67	80	2950	7975	2	19,9
85	30.4	67	80	3250	14275	2	21,8
86	44.6	67	80	1850	3925	3	13,8
87	40.9		80	1835	2575	2	17,3
88	33.8	67	80	2145	3625	3	18
89	29.8	62	80	1845	7000	2	15,3
90	32.7	132	80	2910	8150	3	11,4
91	23.7	100	80	2420	7250	3	12,5
92	35	88	80	2500	5000	2	15,1
93	23.6		80	2905	4250	1	14,3
94	32.4	72	80	2290	4700	3	17
95	27.2	84	81	2490	5100	1	15,7
96	26.6	84	81	2635	5175	1	16,4
97	25.8	92	81	2620	4950	1	14,4
98	23.5	110	81	2725	4550	1	12,6
99	30	84	81	2385	4900	1	12,9
100	39.1	58	81	1755	4400	3	16,9
101	39	64	81	1875	3600	1	16,4
102	35.1	60	81	1760	4000	3	16,1
103	32.3	67	81	2065	3950	3	17,8
104	37	65	81	1975	3775	3	19,4
105	37.7	62	81	2050	4475	3	17,3
106	34.1	68	81	1985	3975	3	16
107	34.7	63	81	2215	3450	1	14,9
108	34.4	65	81	2045	3850	1	16,2
109	29.9	65	81	2380	4100	1	20,7
110	33	74	81	2190	5650	2	14,2
111	34.5		81	2320	4600	2	15,8
112	33.7	75	81	2210	5550	3	14,4
113	32.4	75	81	2350	4650	3	16,8
114	32.9	100	81	2615	5825	3	14,8
115	31.6	74	81	2635	5650	3	18,3
116	28.1	80	81	3230	8500	2	20,4
117		110	81	2800	8225	2	15,4
118	30.7	76	81	3160	8550	2	19,6
119	25.4	116	81	2900	9475	3	12,6
120	24.2	120	81	2930	8375	3	13,8

vehículo	consumo	potencia	año	peso	precio	origen	aceleración
121	22.4	110	81	3415	6675	1	15,8
122	26.6	105	81	3725	6875	1	19
123	20.2	88	81	3060	5450	1	17,1
124	17.6	85	81	3465	5625	1	16,6
125	28	88	82	2605	5275	1	19,6
126	27	88	82	2640	5500	1	18,6
127	34	88	82	2395	5175	1	18
128	31	85	82	2575	5650	1	16,2
129	29	84	82	2525	6250	1	16
130	27	90	82	2735	5650	1	18
131	24	92	82	2865	5225	1	16,4
132	23		82	3035	5150	1	20,5
133	36	74	82	1980	5200	2	15,3
134	37	68	82	2025	4900	3	18,2
135	31	68	82	1970	5100	3	17,6
136	38	63	82	2125	4350	1	14,7
137	36	70	82	2125	4550	1	17,3
138	36	88	82	2160	6350	3	14,5
139	36	75	82	2205	6500	3	14,5
140	34	70	82	2245	5425	3	16,9
141	38	67	82	1965	4625	3	15
142	32	67	82	1965	4875	3	15,7
143	38	67	82	1995	5075	3	16,2
144	25	110	82	2945	7300	1	16,4
145	38	85	82	3015	7450	1	17
146	26	92	82	2585	6000	1	14,5
147	22	112	82	2835	6100	1	14,7
148	32	96	82	2665	7700	3	13,9
149	36	84	82	2370	5475	1	13
150	27	90	82	2950	8400	1	17,3
151	27	86	82	2790	5975	1	15,6
152	44	52	82	2130	5075	2	24,6
153	32	84	82	2295	5650	1	11,6
154	28	79	82	2625	4925	1	18,6
155	31	82	82	2720	6050	1	19,4

- Construya un modelo de regresión para explicar el consumo de los vehículos en función de el peso y la potencia. Indique qué modelo obtiene.
- Verifique si se cumplen las hipótesis del modelo de regresión múltiple ¿detecta alguna anomalía? En caso de detectar alguna anomalía indique que acción correctora adoptaría.
- A continuación estudie la relación entre el precio, como variable dependiente, y la potencia y el peso como variables explicativas.

- d. ¿Cuáles de los parámetros del modelo son significativos con un nivel de significación del 5%?
- e. Los años en que esos modelos salieron al mercado fueron de fuertes crecimientos de los precios. Introduzca el año de salida al mercado del modelo como regresor e indique si es significativo este regresor. Proporcione el nivel crítico del contraste pertinente.
- f. El diagrama de residuos frente a valores predichos muestra cierta curvatura. Tome el logaritmo decimal del precio y repita el análisis. Indique el nuevo valor de R^2 .
- g. El origen de los vehículos podría ser influyente en el precio. Introduzca este factor mediante dos variables cualitativas. Proporcione el modelo final seleccionado.
- h. Construya dos modelos de regresión simple con la aceleración como dependiente y con el peso y la potencia como independientes y después una regresión múltiple con aceleración como dependiente y peso y potencia como independientes. Observe los signos de los coeficientes e interprete lo que ocurre.

6. Se han recopilado los tiempos (dados en segundos) de los campeones olímpicos en las pruebas de 200 metros lisos, 400 metros lisos, 800 metros lisos y 1500 metros lisos de las olimpiadas celebradas entre 1900 y 1996. También se han recopilado los datos referentes a las ciudades en que los Juegos Olímpicos se celebraron y sus altitudes respecto al mar (dadas en metros).

Ciudad	Altitud	fecha	200 m	400 m	800 m	1500 m
París	79	1900	22,20	49,40	121,40	246,00
San Luis	138	1904	21,60	49,20	116,00	245,40
Londres	15	1908	22,40	50,00	112,80	243,40
Estocolmo	15	1912	21,70	48,20	111,90	236,80
Amberes	4	1920	22,00	49,60	113,40	241,80
París	79	1924	21,60	47,60	112,40	233,60
Amsterdam	-2	1928	21,80	47,80	111,80	233,20
Los Ángeles	100	1932	21,20	46,20	109,80	231,20
Berlín	50	1936	20,70	46,50	112,90	227,80
Londres	15	1948	21,10	46,20	109,20	225,20
Helsinki	25	1952	20,70	45,90	109,20	225,20
Melbourne	115	1956	20,60	46,70	107,70	221,20
Roma	15	1960	20,50	44,90	106,30	215,60
Tokyo	14	1964	20,30	45,10	105,10	218,10
México	2220	1968	19,83	43,80	104,30	214,90
Munich	458	1972	20,00	44,66	105,90	216,30
Montreal	53	1976	20,23	44,26	103,50	219,20
Moscú	150	1980	20,19	44,60	105,40	218,40
Los Ángeles	100	1984	19,80	44,27	104,00	212,53
Seúl	34	1988	19,75	43,87	103,45	215,96
Barcelona	0	1992	20,01	43,50	103,66	220,12
Atlanta	320	1996	19,32	43,49	102,58	215,78

- a. Realice una regresión simple para explicar el tiempo empleado por los diferentes campeones olímpicos en función de la distancia ¿es adecuado el modelo? ¿se cumplen las hipótesis del modelo?
- b. Realice una nueva regresión simple en que la variable que se pretende explicar sea la velocidad media en la prueba y el regresor la distancia ¿se cumplen las hipótesis del modelo? sugiera algún modo de remediar los problemas que observe y construya la nueva regresión (simple o múltiple).
- c. Partiendo de los regresores que haya considerado en el punto anterior, construya una nueva regresión (multiple) en que introduzca como nuevos regresores la altitud y el año (o transformaciones de ellas) en que se obtuvieron estas marcas. Estudie posibles modelos alternativos y seleccione el que le parezca más adecuado.

7. Se dispone de los datos de los diámetros, altura y volumen de ejemplares de pino radiata y cerezo negro. Se desea construir un modelo para el volumen utilizando el diámetro y la altura y el tipo de árbol como variables explicativas.

Diam (m)	Alt (m)	Vol (m3)	Z_{pino}	Diam (m)	Alt (m)	Vol (m3)	Z_{pino}
0.21	21.34	0.29	0	0.35	17.71	0.67	1
0.22	19.81	0.29	0	0.05	10.68	0.01	1
0.22	19.20	0.29	0	0.10	12.89	0.05	1
0.27	21.95	0.46	0	0.15	14.38	0.12	1
0.27	24.69	0.53	0	0.20	15.55	0.21	1
0.27	25.30	0.56	0	0.25	16.52	0.33	1
0.28	20.12	0.44	0	0.30	17.36	0.49	1
0.28	22.86	0.52	0	0.35	18.10	0.69	1
0.28	24.38	0.64	0	0.05	13.03	0.02	1
0.28	22.86	0.56	0	0.10	15.72	0.06	1
0.29	24.08	0.69	0	0.15	17.55	0.14	1
0.29	23.16	0.59	0	0.20	18.97	0.26	1
0.29	23.16	0.61	0	0.25	20.16	0.41	1
0.30	21.03	0.60	0	0.30	21.18	0.60	1
0.30	22.86	0.54	0	0.35	22.08	0.84	1
0.33	22.56	0.63	0	0.40	22.89	1.12	1
0.33	25.91	0.96	0	0.10	16.12	0.07	1
0.34	26.21	0.78	0	0.15	18.00	0.15	1
0.35	21.64	0.73	0	0.20	19.46	0.26	1
0.35	19.51	0.71	0	0.25	20.67	0.42	1
0.36	23.77	0.98	0	0.30	21.72	0.62	1
0.36	24.38	0.90	0	0.35	22.64	0.86	1
0.37	22.56	1.03	0	0.40	23.48	1.15	1

Diam (m)	Alt (m)	Vol (m3)	Z_{pino}	Diam (m)	Alt (m)	Vol (m3)	Z_{pino}
0.41	21.95	1.08	0	0.10	17.27	0.07	1
0.41	23.47	1.21	0	0.15	19.27	0.16	1
0.44	24.69	1.57	0	0.20	20.83	0.28	1
0.44	24.99	1.58	0	0.25	22.13	0.45	1
0.45	24.38	1.65	0	0.30	23.25	0.66	1
0.46	24.38	1.46	0	0.35	24.25	0.92	1
0.46	24.38	1.44	0	0.40	25.14	1.23	1
0.52	26.52	2.18	0	0.45	25.96	1.59	1
0.05	7.42	0.01	1	0.15	19.65	0.16	1
0.10	8.96	0.03	1	0.20	21.24	0.29	1
0.15	10.00	0.08	1	0.25	22.57	0.46	1
0.20	10.81	0.15	1	0.30	23.71	0.68	1
0.25	11.48	0.23	1	0.35	24.72	0.94	1
0.05	7.51	0.01	1	0.40	25.63	1.25	1
0.10	9.06	0.04	1	0.45	26.47	1.62	1
0.15	10.11	0.08	1	0.15	21.79	0.18	1
0.20	10.93	0.15	1	0.20	23.56	0.32	1
0.25	11.61	0.24	1	0.25	25.03	0.51	1
0.05	10.45	0.01	1	0.30	26.30	0.75	1
0.10	12.61	0.05	1	0.35	27.42	1.05	1
0.15	14.08	0.11	1	0.40	28.43	1.39	1
0.20	15.22	0.20	1	0.45	29.35	1.80	1
0.25	16.17	0.33	1	0.50	30.20	2.26	1
0.30	16.99	0.48	1	0.55	30.99	2.78	1

8. Se tienen los datos de diferentes países del mundo referidos a las variables alfabetización, población, PIB, tasa de mortalidad infantil.

País	poblacion	alfabetizacion	tasamortinf	PIB
Acerbaján	7400	98	35	3000
Afganistán	20500	29	168	205
Alemania	81200	99	6.5	17539
Arabia Saudí	18000	62	52	6651
Argentina	33900	95	25.6	3408
Armenia	3700	98	27	5000
Australia	17800	100	7.3	16848
Austria	8000	99	6.7	18396
Bahrein	600	77	25	7875
Bangladesh	125000	35	106	202
Barbados	256	99	20.3	6950
Bélgica	10100	99	7.2	17912
Bielorusia	10300	99	19	6500
Bolivia	7900	78	75	730
Bosnia	4600	86	12.7	3098
Botswana	1359	72	39.3	2677
Brasil	156600	81	66	2354
Bulgaria	8900	93	12	3831
Burkina Faso	10000	18	118	357
Burundi	6000	50	105	208
Camboya	10000	35	112	260
Camerún	13100	54	77	993
Canadá	29100	97	6.8	19904
Chile	14000	93	14.6	2591
China	1205200	78	52	377
Colombia	35600	87	28	1538
Corea del Norte	23100	99	27.7	1000
Corea del Sur	45000	96	21.7	6627
Costa Rica	3300	93	11	2031
Croacia	4900	97	8.7	5487
Cuba	11100	94	10.2	1382
Dinamarca	5200	99	6.6	18277
Ecuador	10700	88	39	1085
Egipto	60000	48	76.4	748
El Salvador	5800	73	41	1078
Emiratos Árabes	2800	68	22	14193

País	poblacion	alfabetizacion	tasamortinf	PIB
España	39200	95	6.9	13047
Estados Unidos	260800	97	8.1	23474
Estonia	1600	99	19	6000
Etiopía	55200	24	110	122
Filipinas	69800	90	51	867
Finlandia	5100	100	5.3	15877
Francia	58000	99	6.7	18944
Gabón	1300	61	94	4283
Gambia	959	27	124	351
Georgia	5500	99	23	4500
Gran Bretaña	58400	99	7.2	15974
Grecia	10400	93	8.2	8060
Guatemala	10300	55	57	1342
Haití	6500	53	109	383
Honduras	5600	73	45	1030
Hong Kong	5800	77	5.8	14641
Hungría	10500	99	12.5	5249
India	911600	52	79	275
Indonesia	199700	77	68	681
Irán	65600	54	60	1500
Iraq	19900	60	67	1955
Irlanda	3600	98	7.4	12170
Islandia	263	100	4	17241
Israel	5400	92	8.6	13066
Italia	58100	97	7.6	17500
Japón	125500	99	4.4	19860
Jordania	3961	80	34	1157
Kenia	28200	69	74	323
Kuwait	1800	73	12.5	6818
Letonia	2700	99	21.5	7400
Líbano	3620	80	39.5	1429
Liberia	2900	40	113	409
Libia	5500	64	63	5910
Lituania	3800	99	17	6710
Malasia	19500	78	25.6	2995
Marruecos	28600	50	50	1062

País	poblacion	alfabetizacion	tasamortinf	PIB
México	91800	87	35	3604
Nicaragua	4100	57	52.5	447
Nigeria	98100	51	75	282
Noruega	4300	99	6.3	17755
Nueva Zelanda	3524	99	8.9	14381
Países Bajos	15400	99	6.3	17245
Pakistán	128100	35	101	406
Panamá	2600	88	16.5	2397
Paraguay	5200	90	25.2	1500
Perú	23650	85	54	1107
Polonia	38600	99	13.8	4429
Portugal	10500	85	9.2	9000
Rep. C. Africana	3300	27	137	457
Rep. Dominicana	7800	83	51.5	1034
Ruanda	8400	50	117	292
Rumanía	23400	96	20.3	2702
Rusia	149200	99	27	6680
Senegal	8700	38	76	744
Singapur	2900	88	5.7	14990
Siria	14900	64	43	2436
Somalia	6667	24	126	2126
Sudáfrica	43900	76	47.1	3128
Suecia	8800	99	5.7	16900
Suiza	7000	99	6.2	22384
Tailandia	59400	93	37	1800
Taiwan	20944	91	5.1	7055
Tanzania	29800	46	110	263
Turquía	62200	81	49	3721
Ucrania	51800	97	20.7	2340
Uganda	19800	48	112	325
Uruguay	3200	96	17	3131
Uzbekistán	22600	97	53	1350
Venezuela	20600	88	28	2829
Vietnam	73100	88	46	230
Zambia	9100	73	85	573

Con estos datos se pide responder a las siguientes preguntas:

- Estime el modelo de regresión múltiple para explicar la tasa de mortalidad infantil en función de las restantes variables explicativas. Realice la diagnosis del modelo. ¿Es adecuado? Proporcione indicadores de ello.
- Estime el modelo de regresión múltiple que relaciona la tasa de mortalidad infantil con la alfabetización, el $\log(\text{población})$ y $\log(\text{PIB})$. Interprete los coeficientes de la regresión.

Realice la diagnosis. ¿Se cumplen las hipótesis? Compare los resultados con los obtenidos con el modelo del apartado anterior.

9. Se dispone de los datos de salarios y años de educación de 474 individuos norteamericanos. Se desea estudiar la relación entre ambas variables mediante un modelo de regresión.